AD-A119 781 CHICAGO UNIV IL CENTER FOR DECISION RESEARCH F/G 5/10
A THEORY OF DIAGNOSTIC INFERENCE. II. JUDGING CAUSALITY.(U)
SEP 82 H J EINHORN, R M HOGARTH N00014-81-K-0314

UNCLASSIFIED NL

END
DATE
FILMED
11 .82
DTIC

1

A THEORY OF DIAGNOSTIC INFERENCE:
II. JUDGING CAUSALITY

Hillel J. Einhorn and Robin M. Hogarth
University of Chicago
Graduate School of Business
Center for Decision Research

Sponsored by:
Office of Naval Research
Contract Number, N00014-81-K-0314
Work Unit Number, NR 197-071

September 1982

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER 3	2. GOVT ACCESSION NO. AD-A119781	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) A THEORY OF DIAGNOSTIC INFERENCE: II. JUDGING CAUSALITY		5. TYPE OF REPORT & PERIOD COVERED Technical Report
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Hillel J. Einhorn and Robin M. Hogarth		8. CONTRACT OR GRANT NUMBER(s) N00014-81-K0314
9. PERFORMING ORGANIZATION NAME AND ADDRESS Center for Decision Research, University of Chicago, 1101 East 58th Street, Chicago, Illinois 60637		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR 197-071
11. CONTROLLING OFFICE NAME AND ADDRESS Office of Naval Research 800 North Quincy Street Arlington, VA 22217		12. REPORT DATE August, 1982
		13. NUMBER OF PAGES 70
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Causal judgment, diagnosis, inference, decision making

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Diagnosis involves inferring the process that generated a set of outcomes/results/symptoms. It is essentially causal rather than correlational, looks backward rather than forward, and is constructive in nature. A model of causal reasoning within diagnosis is presented that assumes judgments of causal strength to be a function of the strength of a hypothesis minus the strength of alternatives. However, such judgments are made conditionally on a background or field, analogous to the

over

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-LF-014-6601

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

20.

perception of figure against ground. Report ~~We first~~ demonstrates how changes in the causal background can lead to reversals in judgments of causal strength. The strength of an explanation is then considered as a function of "cues to causality". These are multiple, probabilistic indicators that capture aspects of the data and content inherent in a causal relation. For example, content cues include temporal order, contiguity in time and space, and perceived similarity of cause and effect. Data cues are considered in relation to the concept of perceived validity and the robustness of predictive relations. Literature concerning these cues is summarized and implications regarding the combining of cues are illustrated by the concept of spurious correlation. The importance of imagined alternative explanations is then discussed in relation to answering the counterfactual question: Would Y have occurred if X hadn't ? Experimental evidence is presented concerning use of the cues and the effects of alternative explanations. Results show that people trade-off content and data-based cues to causality and, judged causal strength decreases as a function of the strength of alternatives. ~~Our~~ Results and theoretical model are discussed with respect to the importance of expectations in defining causal relevance, rules for combining the cues to causality, and replacing vs. disconfirming hypotheses. Finally, the normative implications of ~~our~~ the theory are considered with respect to three trade-offs: ~~(1)~~ the acquisition of causal knowledge vs. superstition; ~~(2)~~ "order-out-of-chaos" vs. creativity; and ~~(3)~~ imagination vs. uncertainty.

A Theory of Diagnostic Inference: II. Judging Causality

It was late in middle age that Molière's character, Monsieur Jourdain, made the surprising discovery that he had been speaking prose all his life. Similarly, people may be equally surprised to learn that they have been engaged in diagnostic inference all their lives. By "diagnostic inference" we mean the following: given the occurrence of a set of outcomes/results/symptoms, people infer what causal process could have produced the observed effects. The essential aspects of such inferences are that they are causal rather than correlational, backward rather than forward (one goes from effects to prior causes), concerned with a specific rather than the general case, and constructive (one can synthesize, enlarge, or otherwise develop new hypotheses). The importance of diagnostic inference goes beyond its obvious role in making sense of experience; it is crucial for prediction as well as for defining what is a "relevant" variable (Einhorn & Hogarth, Note 1). Furthermore, since the evidence that one has for making diagnoses is fallible and/or conflicting, the process takes place under uncertainty. Thus, the essential nature of inference, "going beyond the information given" (Bruner, 1957), is as true for diagnosis as for prediction. However, while much attention in the psychological literature has been devoted to prediction (e.g., Kahneman & Tversky, 1973), far less has been given to diagnosis.

In our earlier work (Einhorn & Hogarth, Note 1), a model was developed to describe how people assess the strength of conflicting evidence in diagnostic situations. Specifically, given that some event occurred (e.g., car accident, bank robbery, etc.), we considered the judged strength of particular causal hypotheses based on varying amounts of evidence for each. These judgments were modeled as resulting from an anchoring and adjustment process where the

evidence at hand formed the anchor and adjustments were made on the basis of imagined contrast cases. In fact, the model was interpreted as implying that people combine "what is" with "what might have been" in assessing evidentiary strength. Thus, the ability to imagine alternative scenarios and to engage in counterfactual reasoning was shown to be a crucial component of diagnostic inference, even in simple situations. Moreover, extensions of the model demonstrated how the number and similarity of alternative hypotheses affected the strength of evidence. For example, imagine that there were 10 witnesses to a hit-and-run automobile accident; 4 witnesses reported that the offending car was green and 6 reported blue. On the basis of this evidence, how likely do you think that the offending car was green? Now imagine the situation where 4 reported green, 2 reported blue, 2 reported yellow, and 2 reported red. Would this change your assessment that the car was green? For many people, the different structure of the alternatives implies differences in evidentiary strength. For instance, the fact that four different colors were reported could lead to the diagnosis that the viewing conditions were poor (fog, darkness, etc.), thereby reducing the likelihood of all hypotheses. Similarly, when evidence is missing due to non-responses, it is often seen as diagnostic of certain causal processes. For example, in assessing the likelihood of success of job candidates on the basis of letters of reference, non-responses were deemed to be diagnostic of the writer's unwillingness to say unflattering things about the candidate and were therefore treated as negative evidence.

We interpreted our model and results by drawing an analogy between the processes of diagnosis and perception. In particular: (1) The strength of evidence was seen to be the <u>net</u> effect of the evidence one has compared to the evidence that could have been. Similarly, the salience of objects in percep-

tion can be viewed as the net effect of figure against ground; (2) As in perception, people are constantly engaged in diagnostic inference without active awareness; (3) Diagnosis is a constructive process in that people bring prior expectations to bear in interpreting information and in enlarging hypotheses to account for complex outcomes. In analogous fashion, the importance of expectations and the constructive nature of "achieving" the object are well established in perception (cf. Garner, 1966). Moreover, the introduction of expectations as central to diagnosis (and perception) highlights the role of content knowledge in the assessment of evidence and raises questions of how such knowledge is used.

We take the position that the meaning of evidence is a function, not only of the amount of evidence for or against some position, but of the causal significance that people give to that evidence. Thus, judgments of causality are crucial to diagnosis since the identification of "relevant" variables rests heavily on what is seen as causally relevant, and the judged strength of relations between variables is influenced by the strength of their perceived causal links.

General Approach

We conceive of diagnostic inference as consisting of two inter-related phases: (1) determining what are causally relevant variables; and (2) assessing the strength of causal links. Our focus will be on the second phase although we briefly consider the first. In accord with our basic premise that the strength of evidence involves a figure/ground relation, we propose that the strength of the causal link between some effect Y and a suspected cause X_j $(j = 1, 2, \ldots, n)$, is given by,

$$S_n(Y,X_j|B) = F[s(Y,X_j|B) - \sum_{k=1}^{N} s(Y,X_k|B)] \quad (j \neq k) \tag{1}$$

where,

$S_n(Y,X_j|B)$ = net strength of the causal link of Y with X_j conditional on background B.

$s(Y,X_j|B)$ = gross strength of the causal link of Y with X_j conditional on background B.

$s(Y,X_k|B)$ = gross strength of alternative factor X_k conditional on background B.

F is a monotonically increasing function.

There are several important aspects of equation (1): (a) All terms are conditioned on some assumed or implicit causal background. Thus, the strength of any factor as a cause of Y depends on the context being considered. For example, few persons would attribute the cause of a house fire to the presence of oxygen; however, this could be deemed causally relevant in a fire that took place in a manned space vehicle. Therefore, changes in the assumed causal background can lead to different judgments of a variable's causal strength. We consider this further in the next section; (b) The strength of a causal link is conceptualized as its net strength; i.e., $S_n(Y,X_j|B)$ is a monotonic function of the difference between the gross strength of X_j minus the sum of the gross strengths of specific alternatives. While we examine the determinants of gross strength in detail below, for now it can be viewed as analogous to the absolute strength of a signal perceived against a noiseless background. Thus, the <u>net</u> strength of the $Y \leftarrow X_j$ causal link can be seen as resulting from two conflicting forces; the strength of the signal against the competing signals that comprise imagined alternative explanations. Equation

(1) is therefore a conflict model of evidence that is conceptually identical to our earlier model for assessing evidentiary strength (Einhorn & Hogarth, Note 1). Moreover, it should be noted that equation (1) embodies the "discounting principle" in attribution theory (Kelley, 1973); namely, "the role of a given cause in producing a given effect is discounted if other plausible causes are also present." (p. 113); (c) Equation (1) can be viewed as representing a multi-stage anchoring and adjustment process whereby one anchors on the gross strength of $s(Y,X_j|B)$ and sequentially adjusts for the strengths of imagined alternatives. Indeed, the case for a multi-stage anchoring and adjusting strategy in assessing net strength seems compelling since this decomposes the task into a series of smaller, more manageable components, thereby reducing demands on memory by using a "running total" of causal net strength; (d) Equation (1) posits that net strength follows a difference rather than a ratio model. This has important implications for the case where few or no alternatives are imagined. For example, a ratio model (such as probability theory) would treat the strength of evidence for a hypothesis as certain if there were no alternatives. However, in equation (1), net strength can be low when there are no alternatives if the gross strength of X_j is itself low. Moreover, net strength can also be low when gross strength is high if there are many strong alternatives. Indeed, net strength can only be high if gross strength is high <u>and</u> the strength of specific alternatives is low.

To illustrate these implications, consider the following thought experiment. What is the net strength of any single explanation for the following questions?

1. Why are the outer rings of Saturn braided?
2. Why was Ronald Reagan elected President in 1980?

3. The face of a watch was struck sharply with a hammer.
Why did the glass break and shatter?

For the first question, it is difficult to generate a single explanation, let alone any alternatives, and net strength is low. For the second question, there are many alternatives for any explanation generated (e.g., the situation of the economy; the rise of the "moral majority;" the unresolved hostage crisis; Carter's ineptitude in foreign policy; etc.). Thus, net strength in any particular explanation is also low. On the other hand, for the third question, only one explanation seems plausible; i.e., striking the watch face with a hammer caused the glass to break and shatter. Thus, net strength is high since the explanation is strong (for reasons to be considered later), and it is difficult to imagine specific alternatives. In summary, we argue that like good patterns, good explanations have few alternatives (Garner, 1970); or, to be more precise, whereas good explanations imply few alternatives, the lack of alternatives does not imply good explanations.

Plan of the Paper

Why are some variables deemed to be causally relevant while others are ignored? It is argued that causally relevant variables are perceived as differences in a causal background (Mackie, 1974) and causal strength is a function of that background. An experiment is then described that illustrates the importance of background changes on the judged strength of hypotheses. Thereafter, the mechanism by which the strength of causal links is assessed within a background is examined by discussing the concept of "cues to causality." These cues are fallible indicators of causal relations and are hypothesized to affect the gross strength of explanations. Moreover, the cues are categorized into two classes; those that involve data and those that are

based on content. The trade-off between these categories in judging causal strength is then discussed in relation to the concept of spurious correlation and the fact that causation does not necessarily imply correlation. Following this, the manner in which both types of cues are combined in judgments of causal strength is investigated in several experiments, as is the role of alternative explanations.

The discussion of our experimental results centers on the roles played by the three key concepts of our theoretical framework: (1) the causal background; (2) the cues to causality; and (3) the importance of alternative explanations. Finally, we discuss the normative implications of our work with respect to three trade-offs implicit in causal reasoning.

The Diagnostic Process

Determining Causal Relevance

Much research indicates that processes of perception and judgment are sensitive to differences or deviations from present states, adaptation levels, and reference points (e.g., Helson, 1964; Kahneman & Tversky, 1979). Moreover, we believe that this sensitivity also applies to the types of events that attract attention and arouse diagnostic curiosity. Specifically, we expect that events of diagnostic interest are those that are unusual, abnormal, or unlikely. Thus, one rarely seeks the cause of why one feels "average," why traffic flowed normally, or why some accident is typical. To be sure, diagnostic curiosity can be aroused vis-à-vis normal events. However, we believe this is most likely to happen when those events violate expectations and are therefore seen as unusual after all. For example, we might want to know why traffic flowed normally if major highway improvements were just completed, or why we feel "average" after hearing about a death in

the family. Therefore, diagnostic inference is invoked to make sense of deviations via causal explanation. However, it is important to note that the meaning of a deviation is itself crucially dependent on some assumed background or field. Indeed, even averages can be made unusual with the appropriate shift of background--consider Oscar Wilde's statement that, "moderation shouldn't be taken to extremes."

We propose that in searching for a cause of some outcome which is a deviation from the normal or average, attention is directed toward prior deviations or abnormal events. Thus, unusual effects are seen as the result of unusual causal circumstances. In fact, one can consider this belief a special case of the "representativeness" heuristic (Kahneman & Tversky, 1972) in that causes and effects are similarly discrepant from some assumed causal background. However, the manner in which the causal background affects the strength of causal links needs to be considered in more detail. Specifically, we wish to show that causal relevance is directly related to the degree that a variable is a difference-in-a-background (Mackie, 1974). In order to illustrate this, reconsider our example of oxygen being causally irrelevant in a house fire but relevant in a fire on a spaceship. Furthermore, in accord with equation (1), let Y = fire, X_j = oxygen, B = causal background for the house fire, and, C = causal background for space travel. In the house fire, the gross strength of oxygen as a causal agent is essentially zero since the causal background B already contains the presumption that oxygen was present. Thus, we set $s(Y,X_j|B) = 0$ since X_j is part of B and cannot be a difference in that background. Moreover, if gross strength is zero, net strength will also be zero since there is no need to consider alternative explanations to a worthless hypothesis. This follows from our interpretation of equation (1) as embodying an anchoring and adjustment strategy; i.e., there

is no need for adjustment when gross strength is low. This means that equation (1) only strictly applies when gross strength is sufficiently high so that alternatives are considered. Now consider the spaceship fire; note that $s(Y,X_j|C)$ is not zero since oxygen is not part of the causal background of space flights. Indeed, leaking oxygen would be an important difference-in-the-background. However, this is not to say that oxygen would necessarily be a strong causal candidate since its net strength would depend on alternative explanations. On the other hand, it would not be immediately dismissed as irrelevant, as in the case of a house fire.

Although the causal relevance of a single explanation can be changed by altering contexts, we now demonstrate how the relative likelihoods of two explanations can be reversed by a shift in the causal background. Recall our example in which a watch face is broken (Y) by the force of a hammer (X_1). Let the assumed causal background in the statement of the problem be designated by B, and let X_2 be the alternative explanation that the glass of the watch face was defective. In background B (usual circumstances), we expect the net strength of X_1 to be judged greater than X_2; thus, $S_n(Y,X_1|B) > S_n(Y,X_2|B)$. This follows from equation (1) if it is assumed that $s(Y,X_1|B) > s(Y,X_2|B)$; i.e., the gross strength of the force-of-the-hammer explanation is greater than that of the defective glass. Now consider a change of background to a watch factory (C) where watch faces are tested for defects by striking them with a hammer. In such a background, defects are more of a difference-in-the-background than the force of a hammer, and we expect that $S_n(Y,X_1|C) < S_n(Y,X_2|C)$. Thus, our prediction is that the ordering of the net causal strengths of the two explanations will reverse with changes in the background.

In order to test the above prediction, the following experiment was performed. Subjects (N = 67), recruited from the University of Chicago community, were asked to respond to various experimental stimuli as part of a study on how judgments and choices are made. Each subject was asked to respond to two scenarios, with a gap of some 40 minutes between them (during which time other experimental tasks were administered). Subjects were randomly assigned to one of two groups that received the scenarios in different orders. The two scenarios were as follows:

(1) A watch is placed on a table—face upwards. A hammer is then brought down sharply on the face of the watch. The glass of the watch face breaks and shatters.

(2) In a watch factory, procedures exist for testing various aspects of the end product. One procedure is the following: A watch is placed on a table—face upwards. A hammer is then brought down sharply on the face of the watch. Imagine that on one occasion the glass of the watch face breaks and shatters.

Both scenarios were followed by identical questions:

Question: What caused the glass to break?

a. The force of the hammer?

b. A defect in the glass?

c. Some other explanation (please specify).

Please circle the most likely cause.

Results of the experiment are presented in Table 1, and are shown for the

Insert Table 1 about here

combined groups since the order of presentation had no significant effect. The table shows both the marginal and joint distributions of responses to the different versions of the scenario. In the first situation (background B), 60 subjects (91%) judged the force of the hammer as the most likely cause; however, in the factory setting (background C), the defect in the glass is

seen as the most likely causal agent (36 subjects or 55%). Moreover, of the 60 subjects who said the force of the hammer was the most likely cause in background B, 32 subjects reversed the order in background C. Thus, the prediction that the relative strength of causal hypotheses can be changed by varying the background was substantiated for a majority of our subjects.

In addition to a complete change of backgrounds, equation (1) suggests that net strength can also be altered by narrowing or widening the same background. This occurs because alternative hypotheses are either ruled out by narrowing the context or expanded by widening it. In terms of equation (1), changes in the number of imagined alternatives is represented by different limits of summation in the $\sum_{k=1}^{N} s(Y,X_k|B)$ term. Therefore, the breadth of the background can also affect the causal strength of an explanation. For example, consider the following scenario: Joe is a chemical worker who contracts lung cancer and sues the company for causing his disease. His lawyer argues that the cancer rate of workers in this factory is nine times the national average for workers in comparable industries. Note that the background in this argument is industries of a certain type and the causal argument rests on a difference (higher cancer rates) in this background. However, note how this argument would be strengthened if it could be shown that the cancer rate in this factory was nine times the rate of other chemical factories making exactly the same product. The reason is that by narrowing the field to chemical plants making the same product, the number of alternative explanations is reduced thereby making the difference in the narrowed field more causally relevant. Of course, this does not preclude other classes of alternative explanations. Indeed, the defense lawyer for the chemical company might try to show that the worker has smoked cigarettes for years, comes from a family with a history of respiratory problems, and so on. These

new alternatives clearly reduce the net strength of the chemical-factory-explanation, in accord with equation (1).

The idea that "relevance" can be affected by narrowing or widening the context in which information is given has also been discussed by Bar-Hillel (1980). She considers the research showing that people ignore base rates (prior probabilities) in making probabilistic judgments when individuating information is also present (see e.g., Kahneman & Tversky, 1973). For example, imagine a hit-and-run accident involving a cab and a witness who says the cab was green. The witness is tested and found to be 80% correct in identifying cabs by their color.[1] Moreover, 85% of the cabs in the city are blue while 15% are green. When subjects are asked how likely it is that the offending cab was green, the median and modal response is .80, whereas the Bayesian answer is .41. However, Bar-Hillel states that base rates will be used if they are seen as "relevant" to the question asked. Moreover, she posits that two factors will induce relevance—greater specificity and causal interpretability of the base rate. For example, people will use the base rate in the cab problem if it is stated that 85% of the cab accidents in the city involve blue cabs and 15% involve green (Tversky & Kahneman, 1980); or, the base rate is made more specific, as in stating that, "in the neighborhood in which the accident occurred, which is nearer to the Green Cab Company headquarters than to the Blue Cab Company, 80% of all taxis are Green, and 20% are Blue" (Bar-Hillel, 1980, p. 226). From our perspective, Bar-Hillel's interpretation of relevance is similar to our concept of net strength. Indeed, both are increased by a causal interpretation of evidence and a narrowed background (specificity) which reduces alternatives.

Components of Gross Strength: Cues to Causality

The factors that comprise the gross strength of a causal hypothesis are now considered in detail. Specifically, we hypothesize that gross strength (conditional on an assumed background), is a function of various "cues to causality." The term "cues" has a specific meaning that corresponds with its use in Brunswik's psychology (1952; also see Hammond, 1955; Campbell, 1966). Thus: (1) The relation between each cue and causality is probabilistic. That is, each cue is only a fallible sign of a causal relation; (2) People learn to use multiple cues in making inferences in order to mitigate against the potential errors arising from the use of single cues; (3) The use of multiple cues is facilitated by the intercorrelation (redundancy) between cues in the environment. This both reduces the negative effects of omitting cues, and aids in directing attention to the presence of others; (4) Although multiple cues reduce uncertainty in inference, they do not entirely eliminate it. Moreover, we categorize the cues into two classes; those based on data and those derived from content (cf. Nisbett & Ross, 1980, pp. 97-101). In order to illustrate this distinction, imagine a dichotomized causal candidate X_j, and dichotomized effect Y. Furthermore, consider the four possible outcomes that can result $(X_j \cap Y, \bar{X}_j \cap \bar{Y}, X_j \cap \bar{Y}, \bar{X}_j \cap Y)$. By data-based cues we mean the actual or perceived frequencies of cases in each of the four cells, as well as summary measures of cell combinations (such as statistical correlation). Thus, data-based cues reflect our dependence on general statistical relations in the environment. On the other hand, content-based cues refer to such factors as temporal order of X and Y, contiguity in time and space, and the similarity of cause and effect. These factors depend on one's knowledge of a specific X, Y pair and are not statistical in character. While we discuss both data and content-based cues below, we wish to stress that judgments of gross

strength are a function of both types of cues. Indeed, we will endeavor to show that there are important trade-offs between them.

Data-based cues. The degree to which one variable is believed to be predictive of another, denoted as perceived validity, is an important cue to causality. Note that this cue may be similar or different from the statistical concept of correlation. Indeed, even in statistics, predictive accuracy is not a unitary concept that can be defined without consideration of some loss function (e.g., mean squared error is more appropriate than correlation when predicting exact values of Y). From our perspective, when variables are discrete, perceived validity is a function of 4 cues—the amount of constant conjunction, $(X \cap Y)$; negative constant conjunction, $(\bar{X} \cap \bar{Y})$; and two types of disconfirming data, $(X \cap \bar{Y})$ and $(\bar{X} \cap Y)$. The manner in which people combine these cues in judging covariation has been the subject of much research (see Crocker, 1981 for a review). While the results of individual studies vary, it can be concluded that constant conjunction is the major factor influencing covariation judgments while negative constant conjunction is least important. The role of disconfirming data is less clear; some have argued that people ignore such information while others have shown that negative evidence significantly reduces perceived covariation. Some recent studies speak to this issue.

In a meta-analysis of several covariation studies, Lipe (Note 2) showed that all components of perceived validity, except $(\bar{X} \cap \bar{Y})$, received significant weights when subjects' judgments were regressed on all four data cues. Moreover, Crocker (1982) has found that the form of the question subjects are asked affects the use of disconfirming evidence. Of particular relevance to this issue is a study that addressed how judgments of causal strength are based on the four data cues discussed above (Schustack & Sternberg, 1981).

Specifically, subjects' causal judgments were regressed on all four types of data and results showed significant positive coefficients for both types of constant conjunction, and significant negative coefficients for disconfirming data. Since people can and do use disconfirming data in judging causal strength, by what mechanism does this occur? In accord with our emphasis on the importance of alternatives, we believe that disconfirming data alert people to the existence of alternative explanations without necessarily specifying what they are. Indeed, one source of uncertainty in assessing causal relations may involve a general awareness of one's lack of knowledge of specific causal factors. Thus, our position is that "knowing you don't know" reduces the gross strength of a relation independently of knowledge of specific alternatives.

When variables are continuous, cues to perceived validity include such factors as the rate at which X and Y covary, the direction of the relation, the absolute levels of the variables, and amount of noise (i.e., random error). For example, when subjects are required to learn linear relations in single-cue probability tasks, their performance varies as a function of the magnitude and sign of the slope, the level of the intercept, and the amount of error in the function (Brehmer, 1980). Note, however, that subjects in these tasks are typically not required to judge the strength of relations; rather, they have to learn the function by making predictions and observing outcome feedback. Nevertheless, we view the above factors as cues to relational strength in the continuous case which play an analogous role to the four cues in the dichotomous case. In particular, the level of random error in the former approximates the amount of disconfirming data in the latter. In fact, our assertion that disconfirming data increase general awareness of alternative hypotheses is supported by Brehmer's (1980) results

that subjects search for alternative hypotheses when their predictions are discrepant from outcome feedback (see also, Gettys & Fisher, 1979).

We also postulate that people are sensitive to the extent to which statistical relations vary as a function of other variables and call this cue, the robustness of a relation (see Toda, 1977). For example, imagine that there is a positive correlation between smoking and lung cancer. Now consider that the correlation is computed separately for men and women with the following result: The correlation is positive for men but zero for women. Note that by dividing the original sample into sub-groups, one now considers several relations and asks whether the correlation is the same in all groups. If it is not, the relation is not robust and the causal relevance of smoking for lung cancer is decreased. On the other hand, if a statistical relation is robust, it points more strongly to a causal relation.

The cues discussed above can be conceptualized as forming part of a hierarchical structure. At the most general level, gross strength is a function of data and content; i.e.,

$$s(Y,X_j|B) = f(\text{Data, Content}) \qquad (2)$$

At a second level, the data component of gross strength is a function of perceived validity, $(\Psi[r_{xy}])$, and robustness; thus

$$\text{Data} = g_1(\Psi[r_{xy}], \text{robustness}) \qquad (3)$$

At a more specific level, perceived validity is itself dependent on various cues. That is,

$$\Psi(r_{xy}) = h_1 \begin{cases} (X\cap Y),\ (\bar{X}\cap\bar{Y}),\ (X\cap\bar{Y}),\ (\bar{X}\cap Y);\ \text{in the dichotomous case.} \\ \text{rate, direction, level, noise; in the continuous case.} \end{cases} \tag{4}$$

Therefore, in making causal judgments, specific level cues are combined into higher level units which are themselves cues to the data component of gross strength. Moreover, whereas we have not specified the forms of the functions for combining cues at the different levels, we note that the functions assumed in the studies quoted above are compensatory. For instance, Schustack and Sternberg (1981) used a linear regression model in which judgments of causal strength resulted from the trade-offs between confirming and disconfirming data. Furthermore, one can imagine trade-offs between cues at higher levels; e.g., between perceived validity and robustness. Finally, the fact that trade-offs can occur at all levels of the hierarchy demonstrates the inherent presence of conflict in diagnostic inference (cf. Einhorn & Hogarth, 1981).

Having considered the components of gross strength due to data, we now turn our attention to the components of gross strength due to content.

Content-based cues. We first consider the cue of temporal order, which is used to label which of two variables in a relation is cause and which is effect. The importance of temporal order seems obvious; as, for example, in classical conditioning. Indeed, when the order of presenting the conditioned and unconditioned stimuli is reversed, learning is difficult. Thus, attempts at "backward conditioning" have generally been unsuccessful.

An interesting feature of temporal order is the speed and facility with which it is used—often without explicit awareness. This is particularly the case in the interpretation of language and can be illustrated by contrasting ordinary discourse with a system that is both acausal and atemporal; e.g., probability theory. To illustrate, consider the conjunction "and," which frequently implies temporal order in everyday English (Strawson, 1952); e.g.,

he went into the supermarket and bought some coffee. If "going into the supermarket" and "buying some coffee" are represented by S and K, respectively, how should one understand the question, "What is the probability of S and K?" Whereas a statistician would represent the question as $p(S \cap K)$ and ignore the temporal meaning of "and," others may well perceive the question as formally requiring $p(K|S)$. Indeed, to direct attention to the conjunction of the events, it might be helpful to reverse S and K in order to break the implied time order, i.e., "What is the probability of buying some coffee (K) and going into the supermarket (S)?"

The following experiment demonstrates this point. Forty-eight master of business students (MBA) at the University of Chicago were the subjects. They had all taken at least one graduate level statistics course and were randomly divided into two groups. Subjects in one group were then asked how they would represent, in probabilistic terms, the question, "What is the probability that Joe went into the supermarket (S) and bought some coffee (K)?" The second group was asked how they would represent the question, "What is the probability that Joe bought some coffee (K) and went into the supermarket (S)?" Subjects in both groups were asked to choose their answers from four alternatives: $p(S \cap K)$, $p(S|K)$, $p(K|S)$, none of the above. The results in the first group were that 14 subjects chose $p(S \cap K)$, 9 chose $p(K|S)$, and one said "none of the above" (no one chose $p(S|K)$). In the second group, however, 18 chose $p(S \cap K)$, no one chose $p(K|S)$, 3 chose $p(S|K)$, and 3 picked "none of the above." Thus, in the first group, where "and" is consistent with the usual temporal order of the events considered here, there is much confusion as to whether the statement is one of joint or conditional probability. In the second group there was less confusion although the strong implication of "and" did result in three subjects choosing $p(S|K)$. Tversky

and Kahneman (1980) have also demonstrated the effects of implicit temporal ordering on attempts to interpret the meaning of information presented in probabilistic form.

Our second content cue is contiguity, which is composed of spatial and temporal components. Contiguity is important because it aids in focusing attention on what variables occurred close in time to, and/or in the vicinity of, some effect Y (cf. Michotte, 1946). Indeed, Siegler has shown that for young children (5-6 years old), temporal contiguity is a very strong cue for inferring causality (Siegler & Liebert, 1974; Siegler, 1976). Moreover, these studies show that older children are less dependent on contiguity alone, being able to make use of multiple cues. Nevertheless, in the absence of high contiguity, variables that are causally related may not be noticed as important. For instance, the temporal gap between intercourse and birth requires some knowledge of human biology and chemistry to maintain the links between those events. Similarly, to connect the raising of oil prices in the mid-East with increases in the U.S. inflation rate necessitates an economic model to bridge the spatial gap.

Perceived similarity is a fundamental cue to causality. Indeed, in judgments of gross strength, perceived similarity is to content what perceived validity is to data. Like perceived validity, perceived similarity can be modeled as a function of its elements, some of which add, and some subtract, from its strength. That is, following Tversky (1977), similarity judgments can be defined as a weighted linear function of the common elements of two objects (cf. constant conjunction) minus the distinctive elements of each (cf. disconfirming data). However, to extend this conception of similarity from objects to causes and effects, it is necessary to specify the common and distinctive elements of the latter. These can be considered at several

levels. First, there is a long-standing notion that cause and effect should exhibit some degree of physical resemblance. Mill noted that this is a deeply rooted belief that, "not only reigned supreme in the ancient world, but still possesses almost undisputed dominion over many of the most cultivated minds" (cited in Nisbett & Ross, 1980, p. 115). Mill thought that such a belief was erroneous and many cases exist in which physical resemblance has been misleading. For example, Nisbett and Ross (1980) point out that physical resemblance was the cornerstone of a medical theory called the "doctrine of signatures" whereby cures for diseases were thought to be marked by their resemblance to the symptoms of the disease. Thus, the curing of jaundice was attributed to a substance that had a brilliant yellow color (see also Shapiro, 1960; Shweder, 1977).

At a second level, one can consider similarity based on such elements as the length and strength of cause and effect. That is, if the effect of interest is large (i.e., is of substantial duration and/or magnitude), people will expect the cause(s) to be of comparable size. For example, the germ theory of disease advanced by Pasteur must have seemed incredible to his contemporaries in that people were asked to believe that invisible creatures caused death, plagues, and so on. In the same way, it is equally difficult for many to believe that billions of dollars spent on social programs in the '60s and '70s could have had little or no effect or that long term and complex effects like poverty can have short term and simple causes.

At a higher level of abstraction, elements of similarity judgments may take on metaphorical significance. Thus, in attempting to explain a phenomenon that is not understood, a search is made for one that is, and which has similar features. Deductions from the latter are then extrapolated to the former. Bronowski (1978) has put this well by noting that,

> . . . every act of imagination is the discovery of likeness between two things which were thought unlike. And the example I gave was Newton's thinking of likeness between the thrown apple and the moon sailing majestically in the sky. A most improbable likeness, but one which turned out to be (if you will forgive the phrase) enormously fruitful. (Bronowski, 1978, pp. 109-110).

As was the case for data-based cues, we view content-based cues as also forming part of a hierarchical structure. At a general level, the importance of content in affecting gross strength is a function of temporal order, contiguity, and perceived similarity; i.e.,

$$\text{Content} = g_2[\text{temporal order, contiguity, } \Psi(\text{similarity})] \quad (5)$$

At a second level, contiguity is comprised of temporal and spatial components, while perceived similarity is a function of the common and distinctive elements of cause and effect. Thus,

$$\left.\begin{aligned} \text{Contiguity} &= h_2(\text{time, space}) \\ \Psi(\text{Similarity}) &= h_3(\text{common, distinctive elements}) \end{aligned}\right. \quad (6)$$

Moreover, as discussed above, the elements of perceived similarity can be characterized as relating to physical resemblance, congruity of length and strength, and metaphorical significance.

The existence of conflict among data-based cues is paralleled by conflicts among content-based cues. For example, when temporal contiguity is zero, it nullifies the effect of temporal order since there can be no temporal order for simultaneously occurring events. However, whereas conflicts involving temporal order are usually resolved without difficulty, this is not true of conflicts between perceived similarity and contiguity. Indeed, much evidence exists that is relevant to the conflict between these cues. This evidence can be summarized by considering two levels of each cue (high and

low) in a 2 x 2 table, as shown in Table 2. First consider the high-high

Table 2 about here

cell, examples of which are given by classical conditioning and some of Michotte's (1946) demonstrations of the perception of causality induced by moving objects. In particular, Michotte's subjects reported perceiving causal relations when the movement of objects after contact was congruent with prior trajectories and/or positions. On the other hand, in the lower left hand cell (high contiguity, low similarity), there was no perception of a causal relation. Indeed, Michotte noted that to obtain a causal effect, "requires a certain degree of similarity between the movement of the agent and the change shown in the patient, without which the change would not appear as an 'extension' of the first. This is why there is no causal impression when the movements made by the two objects are in directly opposite directions, or at least very different ones" (Michotte, 1946, p. 210--citation translated by authors). Further evidence concerning this cell is provided by the literature that demonstrates the limits of classical conditioning. For example, whereas Watson and his colleagues were able to condition little Albert to fear rabbits by pairing the appearance of a rabbit with that of a large noise, they could not produce the same effect when the rabbit was replaced by a block of wood or a cloth curtain (Nisbett & Ross, 1980, p. 104). Seligman (1970) reviewed many learning studies and concluded that organisms are differentially prepared to learn different types of relations. The extent to which such limits are biological and could be overcome by experiencing relevant environmental contingencies is, of course, controversial. However, the fact remains that similarity between cause and effect in terms of congruity of length and strength and/or physical resemblance is a crucial cue, and may often be necessary.

Evidence on the comparison between the lower left and upper right cells has been provided by Garcia and his colleagues (Garcia, et al., 1968; Garcia, et al., 1972; also see Garcia, 1981, for a review). They have demonstrated that although rats can learn to associate, after one trial, distinctive tasting food and a gastro-intestinal illness suffered several hours later (induced by x-rays), they do not learn to associate a different shape of food with illness when the taste is familiar. Shultz and Ravinsky (1977) have further investigated the relative weights accorded to similarity vs. temporal contiguity among children varying in age from 6 to 12. For 6-year olds, they found that similarity outweighed temporal contiguity; however, the older children favored contiguity over similarity.

Finally, the conflict between similarity and contiguity can sometimes be resolved by distinguishing between two types of cause. In particular, consider the difference between a "precipitating" and "underlying" cause. The former is generally some action or event that is high in temporal and spatial contiguity but low in similarity of length or strength with the event. The latter is generally based on high similarity of length and strength, with contiguity being less important. Thus, the precipitating cause of World War I was an assassination in Sarajevo, but the underlying cause(s) were economic upheaval, German nationalism, and so on.

Data and Content-Based Cues

In real world judgments, data and content-cues are invariably present, thus raising the question as to how they are combined. We now consider empirical evidence concerning this issue, but viewed within our theoretical framework.

First consider the work on children's causal judgments. Sedlak and Kurtz

(1981) have reviewed the research on factors that influence causal inference and focus on three cues: temporal order, constant conjunction, and spatio-temporal contiguity. Note that these cues are three of the necessary features of the Humean doctrine of causality (Cook & Campbell, 1979); moreover, they involve content and data. In summarizing the results of studies where these cues conflict, Sedlak and Kurtz (1981) have stated:

> When a temporal delay is introduced, only children 8 years of age or older offer reliable evidence of covariation. Specifically, the covariation use of younger children deteriorates when the covarying cause is temporally non-contiguous with the effect and they encounter problems in (1) rationalizing the delay, or (2) recognizing the constancy of the delay. Where these tasks are made difficult, a temporally contiguous (albeit inconsistent) cause is preferred. Thus we can tentatively rank order the three Humean principles. It appears that temporal antecedence supercedes spatial contiguity (Bullock & Gelman, 1979), and that temporal (and spatial) contiguity may sometimes outweigh the covariation factor in children's judgments (Mendelson & Shultz, 1976; Siegler, 1975; Siegler & Liebert, 1974). (Sedlak & Kurtz, 1981, p. 763).

Furthermore, in studies that have contrasted covariation with similarity, Shultz and Ravinsky (1977) found that 6-year olds were unwilling to label dissimilar factors as causes, even in the presence of systematic covariation. On the other hand, older children (10-12 years old) favored covariation over similarity.

In the literature on adults' judgments of covariation, the importance of content is emphasized by the effects of labeling variables. For example, Jennings, Amabile, and Ross (1982) found that when people viewed scatterplots of variables labeled X and Y, the statistical correlation had to be quite high for a relation to be seen. However, when only given the names of pairs of variables, subjects judged the relations to have different strengths. Thus, in the absence of data, the cues evoked by labeling variables affected perceived relational strength. The effects of labeling have also been examined in probability learning studies and confirm the importance of

content-based cues to causality. For example, Adelman (1981) found that subjects in a multiple-cue probability task learned quite well when variable labels were congruent with the strength of the statistical relation but not otherwise. Camerer (Note 3) showed that subjects were able to learn a disordinal interaction only when the variables were labeled in accord with prior beliefs (this involved factors thought to affect the price of wheat futures in a commodity market). When the same task was given as an abstract problem with variables labeled as X_1, X_2, and Y, no learning occurred. In a further probability learning study, Miller (1971) investigated the effects of misleading cue labels. In the presence of such conflict, he found that statistically sophisticated subjects concentrated on the data and ignored the labels, whereas many other subjects adopted the reverse strategy. Thus, in addition to illustrating the conflict between data and content, Miller's (1971) study also demonstrated that differences due to training may affect how such conflicts are resolved.

The conflict between cues that reflect content and data can be highlighted by considering the concept of spurious correlation (Einhorn & Hogarth, 1982). The existence of this concept suggests that some correlations are more (or less) causally related than others, and thereby raises the issue of how to tell the difference (cf. Simon, 1954). For example, consider the correlation between the number of pigs and the amount of pig-iron (Ehrenberg, 1975). Such a correlation seems spurious when the common causal factor, "economic activity," is considered. On the other hand, consider the correlation between amount of rain and number of auto traffic accidents in a city, over the course of a year. Such a correlation does not seem spurious (or at least, less spurious). What is the difference between these two cases?

If we make use of the cues to causality, the spuriousness of the correlation between pigs and pig-iron becomes apparent. That is, although the data point to a statistical relation, the content cues point away from a causal relation; specifically, temporal order (which cannot be used to specify which variable is cause or effect, thereby also nullifying temporal contiguity); low spatial contiguity (it being unlikely that farms and factories are in close physical proximity); and similarity of the variables is only with respect to their names. Indeed, the judgment that the relation is spurious is made easily and quickly.

Now consider the second case: assuming that the data point to a statistical relation, note how the content cues reinforce that link. The temporal order of rain and accidents is clear; contiguity is high both for time and space; and similarity, via the use of prior knowledge about the effects of slippery roads, is high. There seems less doubt that the correlation is "real."

When data and content conflict, spurious correlation is not the only outcome; e.g., a low or zero statistical correlation could mask a true causal relation. To illustrate, imagine that we were ignorant as to the cause of birth. However, it has been suggested that sexual intercourse is related to pregnancy and the following experiment was designed to test this hypothesis: 100 couples were allocated at random to an intercourse condition, and 100 to a non-intercourse condition. As indicated in Table 3, 25 females became pregnant, and 175 did not. In light of our current knowledge (but unknown to our hypothetical selves), we can state that the 5 people in the no-intercourse/yes-pregnancy cell represent "measurement error," i.e., faulty memory in reporting, lying, etc. Since the statistical correlation is small ($r = 0.34$),

<u>Table 3 about here</u>

we might question whether the hypothesis is worth pursuing. Indeed, if the sample size were smaller, the correlation might not even be "significant." Moreover, even with a significant correlation, $r^2 = 0.12$, which is hardly a compelling percentage of the Y variance accounted for by X.

There are two important implications of this example. First, whereas statistics texts correctly remind us that correlation does not necessarily imply causation, the imperfect nature of this cue to causality is also reflected in the statement: <u>causation does not necessarily imply correlation</u>. We have somewhat facetiously labeled examples of the latter as "causalations," giving them equal standing with the better-known and opposite concept of spurious correlation. Second, causalation demonstrates that sole reliance on statistical measures for understanding and interpreting data is insufficient, thereby highlighting the role of judgment (see also Simon, 1954). Moreover, specification of the cues to causality provides some hint as to the components of such judgments.

The Role of Alternative Explanatious

The role of alternative explanations in determining the net strength of a causal relation was explicated in equation (1). Moreover, as stressed by Campbell and colleagues (Campbell & Stanley, 1963; Campbell, 1969; Cook & Campbell, 1979), causal strength should be evaluated by the ruling out of alternative explanations. In fact, Mackie (1974) states that the primitive notion of a cause involves asking oneself the question: "Would Y have occurred if X had not?" The greater the number of alternative explanations underlying a "yes" answer, the lower the causal relevance of X for Y. Note that the posing and answering of the above question (the "counterfactual conditional") may either involve doing a real experiment or what is called a

"thought" experiment. In the former, one compares the effect of X on Y with that of $\bar{X}$ on Y (the control group condition). In this way, the counterfactual question is easily answered. Moreover, note that a control group allows one to infer that X is the only difference in the field that can affect Y (since the control group is the field). However, even in real or quasi-experiments, alternative explanations can exist if other cues to causality are ambiguous. For example, Campbell (1969) points out that: (a) the gradual introduction of some experimental change makes the determination of its causal impact more difficult than a sharp introduction, as for example when a remedial program is phased in over a long time period rather than implemented all at once; (b) unless replication of the $X \cap Y$ relation is accomplished by the introduction of X over multiple time periods or over different units within the same time period, evidence of constant conjunction is weak; (c) the determination of causal relations when effects are not contiguous in time and space with the manipulated variables is problematic. For example, variables that have "lagged" effects or cumulative effects over time are difficult to isolate.

When real or quasi-experiments are not possible, one can nevertheless engage in the following thought experiment in order to answer the question, "Would Y have occurred if X hadn't?": Imagine the world before X, go forward to where X would occur, and then delete it from the scenario. Now run the scenario forward from that point to see if Y happens or not. Clearly, in such thought experiments the construction of "possible worlds" and imaginary scenarios is crucial for judging causal significance (see also Kahneman & Tversky, 1982). In fact, according to Mackie (1974),

> The key item is a picture of what would have happened if things had been otherwise, and this is borrowed from some experience where things were otherwise. It is a contrast case rather than the repetition of like instances that contributes most to our primitive concept of causation. (p. 57)

The use of alternative explanations has important implications for making inferences in general and for interpreting experiments in particular. As recognized by Hume, "not only in philosophy, but even in common life, we may attain the knowledge of a particular cause merely by one experiment, and after a careful removal of all foreign and superfluous circumstances" (as quoted in Mackie, 1974, p. 25). As a case in point, consider the following one-shot case study with a single datum: The occurrence of a huge explosion near Los Alamos, New Mexico, in July 1945. No one doubted this to be the effect of detonating an atomic bomb. Clearly, inferring causality in this poorly designed experiment was not difficult whereas assessing causality in the most meticulously designed experiments in social science is often problematic at best. When one considers why the causal inference is so strong in the bomb example, ask yourself the following question: "Would an explosion of such magnitude have occurred if an atomic bomb had not gone off?" While it is possible to think of alternative explanations for the explosion, they are so unlikely as to be virtually non-existent. Moreover, note how the other cues to causality point to a causal relation (temporal order, contiguity in time and space, similarity of the "unusualness" of effect and cause, and so on). Therefore, even in one-shot case studies with no control group, causality can be inferred (see Campbell, 1975 for an illuminating discussion of this issue).

Tests of the Theoretical Model

Given the various components of our framework, what would constitute an adequate experimental test of its validity? In order to help answer this, we have summarized our theoretical model in the schematic representation shown in Table 4. The overall model is presented as a hierarchical arrangement of cues

Insert Table 4 about here

at different levels of generality. However, the hierarchical structure is not necessary for the model; it is simply a useful way to represent the fact that cues at one level of specificity are combined to form others. Table 4 makes clear that in addressing the question posed above, we are confronted with the issue as to which level or levels to investigate. However, given that we have already reviewed much of the extensive research concerning aspects of levels 1 and 2, we will focus our attention on the less studied aspects represented by levels 3 and 4. In particular, since most everyday judgments involve some mixture of content and data, the real significance of this research is to explicate how these components are combined. Furthermore, given our emphasis on the importance of alternative explanations in affecting causal judgments, we need to test if the conceptual distinction between gross and net causal strength can be empirically demonstrated.

The decision to investigate the roles of content, data, and alternative explanations, has advantages and disadvantages. The chief advantage is that it narrows the empirical work to a manageable level while addressing important issues. However, this comes at the cost of ignoring the many interesting process details suggested by the model. We are sensitive to this concern and can only state that our model invites a programmatic study of causal judgments that we intend to pursue. For now, however, we consider the more general issues.

Experimental Evidence

We now present two experiments that were designed to demonstrate the use of data and content-based cues to causality as well as the role of specific alternative explanations. In both experiments, subjects were first required to read short scenarios and then rate the likelihood that two variables were

causally related. The specific content-based cues varied in this task were contiguity and perceived similarity (defined operationally below); the data-based cues were the four dichotomous data cues on which judgments of perceived validity are based. The role of specific alternatives was investigated by providing subjects with a particular alternative and then asking for a re-assessment of causal strength.

Our hypotheses are: (1) variations in the levels of data and content-based cues will affect net causal strength. Specifically, higher levels of the cues will lead to judgments of stronger causal relations; and, (2) consideration of specific alternatives will reduce judged causal strength. However, such reductions will vary according to the gross strength of the alternatives. These hypotheses can be formalized as follows:

$$\text{H1:} \quad s(Y,X_j|B) = \beta_1[\Psi(r_{xy})] + \beta_2(\text{contiguity}) + \beta_3[\Psi(\text{similarity})] + \varepsilon \qquad (7)$$

Note that we have not included a term to represent interactions since we have no a priori theory regarding the nature of such interactions.

To test for the effects of specific alternatives, note that the judged causal strength of variable X_j can be expressed as,

$$S_{n_1}(Y,X_j|B) = F[s(Y,X_j|B) - v] \qquad (8)$$

where, v represents the gross strength of specific alternatives a subject might generate. However, since we have no control over the subject's use of imagination in generating such alternatives, we assume that this is a constant across subjects. When a specific alternative, X_k, is provided, the subject's re-assessment can be written,

$$S_{n_2}(Y,X_j|B) = F[s(Y,X_j|B) - v - s(Y,X_k|B)] \quad (9)$$

Thus, if we assume that the function F is linear, the difference between the two judgments is,

$$S_{n_1} - S_{n_2} = F[s(Y,X_k|B)] \quad (10)$$

In other words, the difference between the two judgments will be related to the gross strength of the specific alternative considered. In the first experiment, subjects were given alternative explanations all of which had high gross strength. Thus we hypothesize,

$$\text{H2a: } (S_{n_1} - S_{n_2}) > 0 \quad (11)$$

In the second and third experiments, half of the alternative explanations had high gross strength, and half had low gross strength. Thus,

H2b: $(S_{n_1} - S_{n_2})$ is positively related to $s(Y,X_k|B)$.

Experiment 1

Subjects

Thirty-two subjects participated in experiment 1. They were recruited through an advertisement in the University student newspaper and were offered $5 an hour to participate in an experiment on judgment. Their median age was 24, their educational level was high (mean of 4.4 years of post high school education), and there were 16 males and 16 females.

<u>Stimuli</u>

The stimuli consisted of eight scenarios varying in length from 100 to 200 words. These concerned: (1) The efficacy of accounting reports in a chain of supermarkets; (2) the study habits of a graduate student; (3) food-poisoning following a church picnic; (4) weight-loss after attending a health program; (5) the effects of environmental factors on the health of high school students; (6) the playing schedule of a tournament tennis player; (7) the effects of new school textbooks on academic performance; and (8) the relation of diet to the performance of marathon runners. A version of the church picnic scenario is reproduced below.[2]

> A group of 52 young adults is affiliated with a local church. The group was invited to a free picnic last Wednesday, where all the food and beverages were provided by the church's office. Thirty-seven of the members attended. On Sunday many of the members were ill complaining of sore throats and colds. There were 25 group members who were ill; 22 of whom had attended the picnic while 3 had not. Twenty-seven of the members were not ill and of these 15 attended the picnic while 12 had not.
>
> How likely do you think it is that the illness was caused by food poisoning?
>
> <u>NOTE</u> that there may not be a causal relationship.
> (Please mark an X in the appropriate place on the scale below.)

```
|----|----|----|----|----|----|----|----|----|----|
0   10   20   30   40   50   60   70   80   90  100
```

Food poisoning did not cause the illness.	There is a 50/50 chance that food poisoning caused the illness.	Food poisoning was the cause of the illness.

The warning, "<u>NOTE</u> that there may not be a causal relationship," was added to avoid ceiling effects in the use of the 0-100 scale. That is, both our pilot studies and prior research (Peterson, 1980; Lipe, Note 2) indicated

that subjects in an experimental situation tend to assume that there is a relation between two variables if they are asked to assess its strength. Since our interest lay in detecting differences between scenarios involving different levels of the cues, the warning note was added to encourage subjects to use the full range of the scale. The stimulus given above represents a case where contiguity was low (the illness was four days after the picnic), perceived similarity was low (sore throats and colds are not similar to the effects of food poisoning), but the correlation was high (see below). For this scenario, high values for contiguity and perceived similarity were achieved by having the picnic take place on Saturday (i.e., one as opposed to four days before the effect), and having "nausea and stomach cramps" as the symptoms of the illness.

The alternative explanations in experiment 1 were all designed to have high gross strength. For example, the alternative for the above scenario was, "A flu virus was also 'going around' the neighborhood at this time." Thus, for this explanation, both contiguity and perceived similarity were high.

Operational Definitions

Two levels, high and low, of each of the causal cues were made operational in the following manner. For perceived similarity, we first created cause-effect pairs that we deemed to vary in similarity. This was independently verified by having subjects rate the similarity of cause-effect pairs on a 0-10 scale. The mean judgments for the high similarity pairs was 6.7 while the mean for the low similarity pairs was 3.1. Independent ratings were also collected for judgments of similarity for specific alternatives (mean of 7.5). For contiguity, high and low levels were simply defined by their physical values (e.g., time in days). For perceived validity, several

studies have shown this cue to be sensitive to the difference between confirming and disconfirming data (Crocker, 1981; Lipe, Note 2; Schustack & Sternberg, 1981). Thus, in the high condition, the ratio of confirming to disconfirming data was set at approximately 2 to 1; in the low condition, the scenarios contained equivalent amounts of confirming and disconfirming data. In fact, to avoid giving subjects identical patterns of data across scenarios, the distribution of data in the four dichotomous cells was slightly varied. Statistically, the high validity condition can be characterized by correlations between .33 and .40, the low condition by coefficients between .00 and .10.

Procedure and Design

Subjects were presented with a booklet containing the 8 scenarios as well as several other experimental tasks. They were instructed to work at their own pace and the average completion time was 1 hour. The 8 scenarios were interspersed with other material to minimize carry-over effects and to provide variety for the subjects. After reading each scenario, subjects were required to fill out the rating form shown above. Furthermore, they were permitted to make notes or calculations. After completing the rating, they were presented with a specific alternative explanation on the following page. They then re-rated the strength of the original causal variable and proceeded to the next task.

The experiment followed a 4-factor within-subjects design where the first three factors were the causal cues, and the fourth factor contained the 8 scenarios arranged in a Latin-square. Specifically, each subject rated 8 different scenarios, where each scenario contained one of the 2 x 2 x 2 = 8 combinations of the cues. In order to form an 8 x 8 Latin-square, 4 subjects were randomly assigned to each of 8 groups. Since a Latin-square involves an

incomplete design, it should be noted that one cannot test all possible interactions. However, since we have no a priori theory regarding interactions, this is a minor limitation of the design. On the other hand, since some interactions can be tested, we chose to examine those of greatest potential importance.

Results

The results of the experiment are shown in Tables 5 and 6. Consider

Insert Tables 5 and 6 about here

the main effects and interactions of the cues. Table 5 shows that both perceived similarity and validity are significant and in the hypothesized direction (see Table 6). However, there is no effect for contiguity. Moreover, there is a small but significant interaction between perceived similarity and validity such that high levels of both factors have a non-additive effect (see Table 6). Furthermore, there is also a strong scenario main effect as well as two weaker interactions (Table 5). These effects show that the specific content of the scenarios is not only important in itself, but also affects the perception of the causal cues. In particular, the scenario X perceived validity interaction is of theoretical interest since it supports the idea that the effects of data on judged causal strength are influenced by the content in which they occur. Finally, our hypothesis concerning the effects of alternative explanations was strongly confirmed. Specifically, across all scenarios and cue-combinations, the mean judgment of causal strength dropped from 34 to 24 after the introduction of a specific alternative. Furthermore, the reduction was statistically significant within 7 of the 8 scenarios ($p < .01$).

Experiment 2

Rationale

There were three objectives in conducting experiment 2: (1) to replicate the findings of experiment 1; (2) to investigate whether contiguity would be used in scenarios that emphasized this cue more strongly; and, (3) to explore the differential effects of strong vs. weak alternatives on net causal strength.

Subjects

Thirty-two subjects participated in experiment 2. They were recruited in the same way as subjects in experiment 1 and were also paid $5 an hour. Their median age was 22, their educational level was high (mean of 3.3 years past high school), and there were approximately as many males as females.

Stimuli

The stimuli consisted of the first four scenarios used in experiment 1 (see above). However, each scenario was modified to emphasize the contiguity cue. For example, in the picnic scenario described earlier, the timing of potential causes and effects was made salient by introducing a specific event at the picnic (the performance of a comedian) and then emphasizing the time at which certain events occurred relative to this event. In another scenario, two persons were introduced who specifically discussed the issue of timing between the effect and potential cause. At the end of each scenario, subjects were given the same rating scale as in experiment 1.

The alternative explanations in experiment 2 were designed to have either high or low gross strength. The high gross strength alternatives were the same as in experiment 1. The low gross strength alternatives were designed to have low similarity between cause and effect. For example, in the picnic

scenario, the weak alternative was, "The picnic attenders had also participated in a hymn sing immediately after lunch."

Operational Definitions

All cues were defined operationally as in experiment 1. Moreover, the mean rating of similarity for the low gross strength alternatives was 0.9 as compared to 7.5 in the high condition.

Procedure and Design

Experimental procedures were identical to the first experiment. However, the design differed in the following respects: each subject received 4 scenarios rather than 8, but the 8 x 8 Latin-square was maintained by having each pair of subjects complete the 2 x 2 x 2 = 8 combinations of the cues. Thus, we maintained the four-factor, within-subjects Latin-square design as in experiment 1. Furthermore, in order to test for alternative strength, 2 of the 4 scenarios rated by each subject contained strong alternatives and 2 contained weak ones. Weak and strong alternatives were balanced across scenarios and combinations of the causal cues.

Results

The results of experiment 2 are shown in Table 7. As can be seen, the

Insert Table 7 about here

three main effects for the cues are significant, as is the main effect for scenarios. Furthermore, the mean difference in judged causal strength for the high vs. low levels of the cues is quite large; specifically, for perceived validity the means are 47 vs. 30; for perceived similarity the means are 45 vs. 32; for contiguity they are 43 vs. 34. Thus, the results for the first

two cues replicate our findings in experiment 1. However, the significant [redacted] result for contiguity illustrates that this cue can be used in combination with the others if it is made sufficiently salient. Therefore, we consider these results, together with those of experiment 1, as supporting the hypothesis that gross strength can be expressed as a weighted, additive function of both data and content-based cues to causality.

The results for specific alternative explanations are now considered. We first examine each scenario to see if judged causal strength decreased after considering a specific alternative (ignoring whether the alternative had a high or low gross strength). For two of the four scenarios, there was a significant decrease ($p < .01$), while a third was marginal ($p < .07$). The fourth scenario showed no effect (this was the same scenario that showed no effect in experiment 1). However, in comparing the decrease in judged causal strength for the good vs. poor alternatives, we only found a significant difference in mean decrease in one scenario ($p < .05$). Thus, while hypothesis 2a is supported in both experiments, hypothesis 2b is in doubt. In order to investigate this more fully, we performed a third experiment.

Experiment 3

Subjects

Eighty subjects participated in experiment 3. They were all MBA students at the University of Chicago, enrolled in the basic graduate level statistics course.

Stimuli

The stimuli consisted of two of the eight scenarios drawn from those used in the first experiment. The two were characterized as having high levels on all three of the causal cues. The reason for this was to insure that judged causal strength be high enough so as to make the introduction of alternatives relevant; i.e., if the original factor has low causal strength, there is no need for subjects to consider alternative explanations. Furthermore, for each scenario, there were two possible alternatives, one high in gross strength and the other low. Moreover, for each scenario, half of the stimuli were paired with the strong alternative and half with the weak.

Operational Definitions

Since only "high" levels of the cues were used in this experiment, these levels were set as in the previous two experiments. Thus, a 2:1 ratio of confirming to disconfirming data was used for perceived validity, perceived similarity was independently rated as high (mean of 7.4 on a 10 point scale), and contiguity was also high. The similarity ratings for the alternative explanations averaged 8.9 in the strong vs. 1.0 in the weak condition.

Procedure and Design

Subjects were given booklets containing the two scenarios and they were asked to rate the causal strength of a given factor on the same 100 point scale used in the prior experiments. Following this, subjects were given an alternative explanation and then re-rated the causal strength of the original factor. After making these two ratings, the second scenario was considered in the same way. Subjects were randomly assigned to one of two conditions: half the subjects received scenarios paired with strong alternatives, and the other

half received scenarios paired with weak alternatives. In addition, the order of scenario presentation was randomized across subjects.

The design of the experiment followed a 2 × 2 layout with one between-subjects factor (strong vs. weak alternatives) and one within-subjects factor (two scenarios). The dependent variable was the difference between the initial rating and the re-assessment. A significant main effect for strong vs. weak alternatives (in the direction predicted by hypothesis 2b), would mean that the average decrease in the scenarios with strong alternatives is greater than in those containing weak alternatives.

Results

The results of the analysis-of-variance are shown in Table 8. The major

Insert Table 8 about here

finding is that the main effect for strong vs. weak alternatives is highly significant. Indeed, the average decrease for strong alternatives is 19.2 (over both scenarios), while the average decrease for weak alternatives is only 1.8. Thus, hypothesis 2b is strongly supported by these data.

Discussion

This paper has emphasized the fundamental role of causal judgments in diagnostic inference. Furthermore, we have demonstrated three crucial aspects of such judgments: (1) that causal judgments are made in relation to a causal background or field; (2) that people use multiple, probabilistic cues to causality in forming their judgments; and (3) that the strength and specificity of alternatives affects judged causal stength. Moreover, these ideas can be summarized by a perceptual analogy in which figures are seen against ground (causal candidates are differences-in-a-background), good

figures are consistent with Gestalt principles (good explanations arise from internally consistent patterns of cues), and, good figures have few alternatives (as do good explanations). Moreover, these ideas are embodied in equation (1), which is reproduced here for convenience.

$$S_n(Y,X_j|B) = F[s(Y,X_j|B) - \sum_{k=1}^{N} s(Y,X_k|B)] \quad (j \neq k) \qquad (1)$$

Our empirical results and their implications are now discussed in light of equation (1).

Causal Background (B)

Causal judgments are made conditionally on some background and causal candidates are seen as differences-in-that-background. However, the issue arises as to how the background or field is defined in any given situation. One important factor concerns the expectations a person brings to the inferential task. These expectations, in turn, are based on assumptions that one may be unaware of. Consider our experiment concerning the breaking of a watch face glass by a hammer. When subjects were presented with this stimulus, they assumed a normal setting in which the breaking of the glass was attributed to the force of a hammer. However, when the context was made explicit, as when watch faces are tested in a factory, the majority of subjects indicated that the cause was due to a defect in the glass. Therefore, in the former, one expects the force of a hammer to be sufficient for breaking the glass, while in the latter, one doesn't expect owners of watch factories to employ testing procedures that are financially ruinous. Thus, defects in the glass are seen as being of greater causal relevance.

The idea that assumptions and expectations can affect behavior is well documented in the psychological literature. Recall Bruner and Postman's (1949) classic experiment demonstrating that subjects shown playing cards with a red six of spades, for example, reported seeing a purple six of hearts or purple six of spades. More recently, Loftus (1979) has shown that eyewitnesses often report what seem to be compromises between what they actually saw and the presupposition underlying the question they are asked. Thus, "leading questions" seem to work by imposing an implicit background on the unsuspecting witness such that the strength of evidence is changed. The relevance of these findings for causal judgments is direct; the causal background against which judgments are made is crucial for assessing causal strength. Moreover, the fact that shifts in background can and do occur, coupled with lack of awareness of the background, suggest that disagreement and controversy will be endemic in discussions of causality. Indeed, the search for an agreed on definition of cause, based solely on structural aspects of inferential tasks, would seem difficult given the essential role of context in defining and assessing causal relations. Therefore, as emphasized in equation (1), causal strength is a relation between factors, and not a thing-in-itself.

A second aspect concerning background effects involves the narrowing or broadening of the background. For example, consider the issue of reductionism in causal explanations, where causes at a molar level are different from those at a molecular level. If we consider the background B as analogous to the field of vision under a microscope, then shifts in magnification of the lens define different fields. Moreover, since causal relevance is a difference-in-the-field, it is obvious that a cause at one level will not necessarily be relevant at another. This microscope analogy makes clear that the "appropriate" level of magnification depends on one's purposes and the extent

of one's knowledge of the phenomenon in question. Thus, a biochemist may see the causal link between smoking and lung cancer as due to chemical effects of tar, nicotine, and the like, on cell structure, while an immunologist might see the causal link as due to the suppression of the immune system in controlling diseases in general. However, it should be noted that the level of the field is not totally arbitrary in everyday inferences. Indeed, there is remarkable consensus among individuals as to the appropriate level of the assumed background. On the other hand, where large discrepancies exist in knowledge about a particular topic, as in comparing experts to non-experts, such consensus is often lacking.

Finally, when expectations that rest on an assumed background are violated, this can be an important cue for reorganizing or re-structuring one's hypotheses. For example, imagine a hit-and-run accident in which all 10 witnesses said the offending car was going 73 miles per hour at the moment of impact. Since we expect much greater variability in such estimates, as well as round numbers, this surprising unanimity might cue one to ask whether the witnesses had colluded in their responses. Similarly, in our earlier paper (Einhorn & Hogarth, Note 1), we noted that the structure of outcomes can suggest new hypotheses such that the diagnosis contradicts the surface meaning of the evidence. Thus, scientific data that are too perfect can suggest fraud (see, for example, Kamin, 1974, on Burt's twin data; Bishop, et al., 1975, on Mendel's pea experiments), evidence in a trial that is too consistent and obvious can suggest the defendant was "framed," and one can "protesteth too much" in a variety of circumstances. Such examples illustrate that violations of expectations can trigger a re-structuring of alternatives. Of course, specifying the conditions that lead to re-structuring as opposed to other responses remains an important, and unanswered question.

Gross Strength and the Cues to Causality

The results of experiments 1 and 2 show that judgments of causal strength are affected by the causal cues of perceived similarity and validity, as well as by contiguity. While the latter was not significant in the first experiment, increasing its salience in the second experiment did lead to its use. Of course, we did not test all of the cues previously discussed since this would require many studies. However, in reviewing the psychological literature in various sub-fields, we pointed to several studies that have varied particular cues, with results that are consistent with our framework. Moreover, whereas we made no claim that the particular cues that we considered are new, we do claim that the conception of these factors as probabilistic cues (in the Brunwikian sense), and their categorization as "content" or "data" based variables, is new. Furthermore, by considering causal judgments as resulting from the weighting and combining of cues, "causality" can be discussed as a psychological topic within the well-defined and researched area of behavioral decision theory, thereby directing attention to issues that might otherwise be ignored. We now consider some of these: (a) What are the ecological validities of the cues?; (b) What role does cue redundancy (intercorrelation) play in causal judgments?; (c) What are the function forms that relate lower level to higher level cues?

(a) It has been assumed throughout this paper that the cues to causality have imperfect but non-zero ecological validities; i.e., each cue is predictive of a true causal relation. How do we know this? Simply put, we don't. The reason is that without some measure of "true" causality, no determination of accurate causal knowledge is strictly possible. However, the fact that the cues we have considered are implicated in a wide variety of studies with both human and animal subjects, leads us to believe that they would not continue to

be used if they were useless. Therefore, our argument is a functional and practical one; viz., given the importance of learning and inferring causal relations for survival, we would find it surprising that the cues on which this depends are totally worthless. On the other hand, we do not advocate the position that if something is used, it must be beneficial for the organism. Such a view is untenable for many reasons (see Einhorn & Hogarth, 1981). Thus, while we have assumed that the cues to causality have ecological validity, their imperfect validity can also lead to superstitions, myth, magic, and the like. Indeed, the relation between causal knowledge and superstition is discussed in the next section.

(b) While we have treated the cues to causality as conceptually distinct, it seems likely that they are correlated in the environment. However, the determination of these correlations would require an elaborate (and problematic) ecological analysis that is beyond the scope of this paper. Nevertheless, the assumption of correlated cues seems warranted since people have strong expectations concerning what cues go together. Indeed, just as in the perception of incomplete figures (where one fills in the missing parts), scenarios are filled in by assuming that cues not explicitly mentioned are in fact present. Thus, the fact that one generally perceives the world as coherent, suggests that the cues to causality are redundant to some degree. Given this, an important implication of redundancy concerns its beneficial effects with regard to simplifying information search and the combining of cues. For example, when cues are redundant,

> (a) Information search is limited without large losses in predictive accuracy; (b) attention is highly selective; (c) dimensionality of the information space is reduced, thereby preventing information overload; (d) intersubstitutability of cues is facilitated (Hammond, 1972); and (e) unreliability of cues is alleviated by having multiple measures of the same cue variable (Einhorn, Kleinmuntz, & Kleinmuntz, 1979, p. 466).

Therefore, although we have posited several causal cues, it is unlikely that all will be used or attended to in every situation. Indeed, recall that the contiguity cue was not used in the first experiment, but was used when attention was directed to it in experiment 2. Moreover, while correlated cues allow one to fill in missing information in making inferences, their less than perfect redundancy implies that in "going beyond the information given," some error is inevitable. However, the fact that inferences are often accurate attests to the prevalence and functional value of redundant cues. (c) In considering how cues are combined at various levels, recall Table 4, which provides a schematic representation of our theoretical model. A question that is common to all levels is the exact nature of the functions that relate components on the right-hand side of the equations to those on the left. In fact, this general question is the focus of much attention in the judgment and choice literature (e.g., Anderson, 1979; Edwards, 1977; Hammond, McClelland & Mumpower, 1980). However, rather than considering each function in Table 4, we consider the larger issue as to whether the rules for combining causal cues are compensatory, noncompensatory, or some combination of both. That is, do causal cues trade-off or not, and under what conditions?

In our experimental work, and in all of the research we have quoted, a compensatory rule (usually additive) was assumed for combining cues into overall judgments of causal strength. As in much other judgment research, such rules have been found to work remarkably well (Einhorn, et al., 1979). Nevertheless, there are reasons to suspect that certain causal cues may not trade-off at all (e.g., temporal order) or, trade-offs will only occur beyond some threshold or minimal level. In particular, the cue of perceived similarity is likely to be one in which some amount must be present before any causal relation is seen. However, given that this level is maintained, trade-

offs of perceived similarity with other cues can occur (as demonstrated in experiments 1 and 2). Thus, some combination of noncompensatory and compensatory rules is likely to be used in conjunction with this cue (cf. Luce, 1978; Payne, 1976). Indeed, we expect that this pattern is generally true of the content-based cues.

The Role of Specific Alternatives

Following the prediction implicit in equation (1), the provision of explicit alternatives with high gross strength significantly reduced judged causal strength in experiments 1 and 2. Moreover, the differential decrease for strong vs. weak alternatives was unequivocally demonstrated in experiment 3. These results are consistent with findings from other studies. For example, in the previously mentioned study by Schustack and Sternberg (1981), subjects considered data relating to alternative hypotheses. A regression analysis of subjects' causal judgments showed a significant negative weight for the strength of the alternatives, thus paralleling equation (1). In a related vein, Koriat, Lichtenstein, and Fischhoff (1980) found that when subjects generated specific arguments against a position they adopted, their confidence in the original position was reduced.

It was stated earlier (p.8), that we viewed equation (1) as embodying an anchoring and adjustment strategy. Furthermore, we stipulated that when gross strength (the anchor) is low, there would be no need to adjust for alternative explanations. After all, why waste effort on a worthless explanation? Data from experiment 1 speak to this point. Subjects in the high perceived similarity- high perceived validity condition (over all eight scenarios) gave an initial mean rating of 57 for causal strength, as opposed to 18 for the

versions containing low levels of these cues. However, after seeing the same strong alternative, the former judgments decreased to 37 while the latter only decreased to 15 (i.e., a decrease of 20 vs. 3). Thus, strong alternatives "hurt" strong explanations more that they hurt weak ones. While this is not surprising on an absolute basis, note that the proportional drop is much greater for the good as opposed to the poor original explanations (35% vs. 18%). This result is suggestive that our hypothesis given above is correct. It is not, however, definitive since we forced people to consider a specific alternative. In a natural setting, we would expect that the gross strength of a causal candidate has to reach some threshold value before people engage in counterfactual reasoning.

The notion of a gross strength threshold emphasizes the need to specify the form of equation (1) more precisely, as well as the scale properties of the gross strength variable. For example, we have assumed that the gross strengths of alternatives in our experiments were non-negative and indeed, the introduction of specific alternatives generally led to a decrease in judged causal strength. However, explicit alternatives could have negative gross strength (by considering them in relation to the background, B), which would lead to an increase in the net strength of a causal candidate. For example, imagine that you believe strongly in a scientific theory for which there are few competitors. You are then presented with an absurd alternative explanation which leads to the following inference: if this is the best alternative that people can generate, your belief in the original theory should be increased. While the frequency of such inferences is not known, equation (1) can handle these reversals of net strength at a conceptual level. However, it can do no more that this without a better measure of gross strength.

Whatever the merits of considering a gross strength threshold, the purpose of diagnostic inference is to generate some causal explanation for observed effects. Thus, while a particular explanation may be poor and not require specific alternatives for contrast, this does not mean that the process terminates at this point. Indeed, one is still left with the question, "If it wasn't X, what did cause Y?" Therefore, while the testing of hypotheses via comparison with alternatives is part of diagnostic inference, it also involves a continuing search for better explanations. The distinction between testing hypotheses and searching for better ones can be likened to a "disconfirmation" vs. "replacement" model of inference. As we pointed out before (Einhorn & Hogarth, Note 1), the replacement model,

> ....is consistent with the Kuhnian notion that theories in science are not discarded, despite evidence to the contrary, if they are not replaced by better alternatives (Kuhn, 1962)...We believe that the replacement view is equally strong in everyday inference...Unless disconfirmation leads to replacement, it can only increase uncertainty and anxiety. A useful analogy might be the following: how many people would read detective stories if the author only revealed who didn't do it? (p. 46)

Furthermore, note that replacement is more powerful than disconfirmation since the former can subsume the latter. For example, compare most real trials, when the accused is found innocent because of reasonable doubt vs. trials depicted in movies (e.g. Perry Mason), when the accused is found innocent by unmasking the real culprit. Such demonstrations simultaneously disconfirm the hypothesis of guilt for the accused, while answering the question: if not X, then who?

Normative Issues

Our theoretical and experimental findings raise the following question: what should be the role of content and imagination in diagnostic inference? In particular, since there is no agreed-on theory of causality, is it possible

to say anything about the quality of diagnostic inferences and the causal judgments on which they are based? While we have no definitive answers, we discuss three trade-offs that are germane to this question: (1) acquisition of causal knowledge vs. superstition; (2) achieving "order-out-of- chaos" vs. limiting creativity; (3) using imagination vs. increasing uncertainty.

Causal knowledge vs. superstition

We have asserted that the cues to causality have ecological validity and that accurate causal knowledge depends partly on their use. However, since the cues are imperfectly valid, the discovery of causal relations can be likened to a complex, multivariate signal detection task where the presence of cause is sought against a background of randomness or noise (cf. Lopes, in press). There are several implications of this signal detection analogy. First, people must set a cut-off point to decide whether or not some factor is to be considered a cause. Second, the position of this cut-off will reflect two types of errors and their associated costs. That is, on the one hand, people can infer causes when they do not exist; on the other hand, they can make the error of failing to infer true causal relations. Moreover, whereas several studies have addressed the former and discussed human susceptibility to "illusions of control" (e.g., Langer, 1975), there has been less awareness of illusions of lack of control (however, see Seligman, 1975; Alloy & Abramson, 1979). Nonetheless, given the importance of inferring causal relations for survival, one could argue that the former illusion is less costly than the latter. Indeed, one can consider superstition as the price that one pays for causal knowledge (cf. Skinner, 1966), although it is an open question as to whether the price is worth the benefits in any particular situation. Finally, a third implication of the signal detection analogy is

that people may exhibit differential sensitivity to seeing causal relations through either training (e.g., developing expertise), or ability.

In addition to using the cues to causality, our position also implies that accurate causal judgment involves the elimination of alternative explanations. However, as vividly demonstrated by the concept of spurious correlation, several variables may be highly correlated in the natural ecology so that the determination of causal relations is problematic. Thus, from this viewpoint one can sympathize with the ingenuous teenager who asked Dear Abby: would she get pregnant from holding hands with her boy-friend? Given that this causal candidate and the true cause are both correlated and share many of the same cues to causality, only a true experiment could resolve the issue. Indeed, the importance of experiments for disentangling correlated factors has been stressed by Hammond (1978). He points out that much learning through experience often rests on the weakest mode of inference--unaided judgment based on passive observation. From a normative viewpoint, the prevalence of correlated alternatives reinforces the need for experimentation in making valid causal inferences (cf. Einhorn & Hogarth, 1978).

Order out-of-chaos vs. creative thought

The causal field and the cues to causality both play an important role in limiting the number of interpretations people make in inferential tasks, and thus in creating "order-out-of-chaos". Furthermore, the adoption of a particular background and the use of the cues proceed quickly and are often marked by a lack of awareness that a delimiting process has taken place. The benefits to be gained from such automatized processes are large. However, they come at the cost of reducing the probability that people can achieve more creative representations of inferential tasks. Indeed, Campbell (1960) has

stressed the importance of deliberately introducing random variation to stimulate creative efforts, especially in science. Without such random perturbations, he argues that the forces that maintain a person's particular conception of a problem are too strong. Moreover, the literature on creativity has many examples of techniques that are aimed precisely at making people aware of the delimiting assumptions they bring to tasks (see, e.g., Adams, 1976). In addition, when using such techniques, people are often requested to refrain from counterfactual reasoning and to make specific use of analogies and paradox to enjoin previously disconnected ideas. In short, to restructure problems in creative ways frequently requires attempts to counter the habitual forces of causal reasoning.

Using imagination vs. increasing uncertainty

Our position is that imagination should play an important role in diagnostic inference. As noted above, imagination is clearly implicated in creative thought. Moreover, we find much merit in using imagination via an anchoring and adjustment process in the testing of hypotheses. That is, we believe that comparisons of "what is" with "what might have been" have many benefits. First, whereas the generation of specific alternatives to a given hypothesis increases uncertainty in that hypothesis, such uncertainty can be beneficial for counteracting overconfidence in judgment. Indeed, it has been argued that one of the most prevalent cognitive biases is "cognitive myopia," which results precisely from the failure to imagine alternative futures (Hogarth, 1981). Second, whereas one can disconfirm a hypothesis by observing data, imagination is necessary for the more important task of replacement (see above).

The use of imagination requires considerable mental effort and is not,

therefore, costless. Moreover, by imagining the world as it might have been, increased uncertainty may lead to feelings of anxiety, an observation which leads us to speculate that imagination may provide a crucial link between cognition and affect (cf. Zajonc, 1980). That is, whereas we have limited ourselves to cognitive comparisons between "what is" and "what might have been", such a process may also underlie various emotional reactions. For example, one might consider regret as the difference between what I did and what I might have done (Kahneman & Tversky, 1982); guilt as the difference between what I did and what I should have done; shame as what I did vs. what others expected me to have done; equity as what I have vs. what others have; and so on. Clearly, much work needs to be done before this suggestion can be taken seriously. However, we have at least provided an anchor for further research on this issue.

Conclusion

We began this paper by emphasizing that diagnostic inference involves judgments of causality made under conditions of uncertainty. Moreover, such judgments were formalized in a model consisting of three main components: causal background, cues to causality, and specific alternatives. Whereas our model accounts for many findings in the literature as well as our own experimental results, it by no means explicates all aspects of causal reasoning. In particular, judgments of causation in the social realm (e.g., attribution theory) were not explored although many of our ideas may prove germane to this topic. However, given the complexity of causal judgment, it seems appropriate to have started with a simple model based on alternatives, background, and cues; i.e., the ABC of causality.

Footnotes

This work was supported by a contract from the Office of Naval Research. We would like to thank the following people for their assistance on this project: Zvi Gilula, Marlys Lipe, Haim Mano, Ann McGill, and Werner Wothke.

[1]This means that p(report of blue|cab was blue) = p(report of green|cab was green) = .8.

[2]Copies of all scenarios can be obtained from the authors.

Reference Notes

1. Einhorn, H.J., & Hogarth, R.M. A theory of diagnostic inference: I. Imagination and the psychophysics of evidence. Center for Decision Research, Graduate School of Business, University of Chicago, June 1982.
2. Lipe, M.G. A cross-study analysis of covariation judgments. Center for Decision Research, Graduate School of Business, University of Chicago, April 1982.
3. Camerer, C.F. The validity and utility of expert judgment. Unpublished Ph. D. dissertation, University of Chicago, 1982.

References

Adams, J.L. Conceptual blockbusting: A pleasurable guide to better problem solving. San Francisco: San Francisco Book Co., 1976.

Adelman, L. The influence of formal, substantive, and contextual task properties on the relative effectiveness of feedback in multiple-cue probability tasks. Organizational Behavior and Human Performance, 1981, 27, 423-442.

Alloy, L.B., & Abramson, L.Y. Judgment of contingency in depressed and nondepressed students: Sadder but wiser? Journal of Experimental Psychology: General, 1979, 108, 441-485.

Anderson, N.H. Algebraic rules in psychological measurement. American Scientist, 1979, 67, 555-563.

Bar-Hillel, M. The base-rate fallacy in probability judgments. Acta Psychologica, 1980, 44, 211-233.

Bishop, Y.M.M., Fienberg, S.E., & Holland, P.W. Discrete multivariate analysis: Theory and practice. Cambridge, MA: The MIT Press, 1975.

Brehmer, B. In one word: not from experience. Acta Psychologica, 1980, 45, 223-241.

Bronowski, J. The origins of knowledge and imagination. New Haven: Yale University Press, 1978.

Bruner, J.S. Going beyond the information given. In J.S. Bruner et al. (Eds.), Contemporary approaches to cognition. Cambridge, MA: Harvard University Press, 1957.

Bruner, J.S., & Postman, L. On the perception of incongruity: A paradigm. Journal of Personality, 1949, 18, 206-223.

Brunswik, E. The conceptual framework of psychology. Chicago: University of Chicago Press, 1952.

Campbell, D.T. Blind variation and selective retention in creative thought as in other knowledge processes. Psychological Review, 1960, 67, 380-400.

Campbell, D.T. Pattern matching as an essential in distal knowing. In K.R. Hammond (Ed.), The psychology of Egon Brunswik. New York: Holt, Rinehart and Winston, 1966.

Campbell, D.T. Reforms as experiments. American Psychologist, 1969, 24, 409-429.

Campbell, D.T. Degrees of freedom and the case study. Comparative Political Studies, 1975, 8, 178-193.

Campbell, D.T., & Stanley, J.C. Experimental and quasi-experimental designs for research. Chicago: Rand-McNally, 1963.

Cook, T.D., & Campbell, D.T. Quasi-experimentation: Design and analysis for field settings. Chicago: Rand-McNally, 1979.

Crocker, J. Judgment of covariation by social perceivers. Psychological Bulletin, 1981, 90, 272-292.

Crocker, J. Biased questions in judgment of covariation studies. Personality and Social Psychology Bulletin, 1982, 8, 214-220.

Edwards, W. Use of multiattribute utility measurement for social decision making. In D.E. Bell, R.L. Keeney, & H. Raiffa (Eds.), Conflicting objectives in decisions. New York: Wiley, 1977.

Ehrenberg, A.S.C. Data reduction: Analyzing and interpreting statistical data. New York: Wiley, 1975.

Einhorn, H.J., & Hogarth, R.M. Confidence in judgment: Persistence of the illusion of validity. Psychological Review, 1978, 85, 395-416.

Einhorn, H.J., & Hogarth, R.M. Behavioral decision theory: Processes of judgment and choice. Annual Review of Psychology, 1981, 32, 53-88.

Einhorn, H.J., & Hogarth, R.M. Prediction, diagnosis, and causal thinking in forecasting. Journal of Forecasting, 1982, 1, 23-36.

Einhorn, H.J., Kleinmuntz, D.N., & Kleinmuntz, B. Linear regression and process-tracing models of judgment. Psychological Review, 1979, 86, 465-485.

Garcia, J. Tilting at the paper mills of academe. American Psychologist, 1981, 36, 149-158.

Garcia, J., McGowan, B., Ervin, F.R., & Koelling, R. Cues: Their relative effectiveness as reinforcers. Science, 1968, 160, 794-795.

Garcia, J., McGowan, B., & Green, K. Sensory quality and integration: Constraints on conditioning. In A.H. Black & W.F. Prokasy (Eds.), Classical conditioning II: Current research and theory. New York: Appleton-Century-Crofts, 1972.

Garner, W.R. To perceive is to know. American Psychologist, 1966, 21, 11-19.

Garner, W.R. Good patterns have few alternatives. American Scientist, 1970, 58, 34-42.

Gettys, C. F., & Fisher, S. D. Hypothesis plausibility and hypothesis generation. Organizational Behavior and Human Performance, 1979, 24, 93-110.

Hammond, K.R. Probabilistic functionalism and the clinical method. Psychological Review, 1955, 62, 255-262.

Hammond, K.R. Toward increasing competence of thought in public policy formation. In K.R. Hammond (Ed.), Judgment and decision in public policy formation. Denver: Westview Press, 1978.

Hammond, K.R., McClelland, G.H., & Mumpower, J. Human judment and decision making. New York: Praeger, 1980.

Helson, H. Adaptation-level theory. New York: Harper, 1964.

Hogarth, R.M. Beyond discrete biases: Functional and dysfunctional aspects of judgmental heuristics. Psychological Bulletin, 1981, 90, 197-217.

Jennings, D., Amabile, T.M., & Ross, L.D. Informal covariation assessment: Data-based vs. theory-based judgments. In D. Kahneman, P. Slovic, & A. Tversky (Eds.), Judgment under uncertainty: Heuristics and biases. New York: Cambridge University Press, 1982.

Kahneman, D., & Tversky, A. Subjective probability: A judgment of representativeness. Cognitive Psychology, 1972, 3, 430-454.

Kahneman, D., & Tversky, A. On the psychology of prediction. Psychological Review, 1973, 80, 251-273.

Kahneman, D., & Tversky, A. Prospect theory: An analysis of decision under risk. Econometrica, 1979, 47, 263-291.

Kahneman, D., & Tversky, A. The simulation heuristic. In D. Kahneman, P. Slovic & A. Tversky (Eds.), Judgment under uncertainty: Heuristics and biases. New York: Cambridge University Press, 1982.

Kamin, L.J. The science and politics of IQ. Potomac, MD: Erlbaum, 1974.

Kelley, H.H. The processes of causal attribution. American Psychologist, 1973, 28, 107-128.

Koriat, A., Lichtenstein, S., & Fischhoff, B. Reasons for confidence. Journal of Experimental Psychology: Human Learning and Memory, 1980, 6, 107-118.

Langer, E.J. The illusion of control. Journal of Personality and Social Psychology, 1975, 32, 311-328.

Loftus, E.F. _Eyewitness testimony_. Cambridge MA: Harvard University Press, 1979.

Lopes, L.L. Doing the impossible: A note on induction and the experience of randomness. _Journal of Experimental Psychology: Human Learning, Memory, and Cognition_, 1982, in press.

Luce, R.D. Lexicographic tradeoff structures. _Theory and Decision_, 1978, _9_, 187-193.

Mackie, J.L. _The cement of the universe: A study of causation_. Oxford: Clarendon Press, 1974.

Michotte, A. _La perception de la causalité_. Paris: Vrin, 1946.

Miller, P. McC. Do labels mislead? A multiple cue study, within the framework of Brunswik's probabilistic functionalism. _Organizational Behavior and Human Performance_, 1971, _6_, 480-500.

Nisbett, R.E., & Ross, L.D. _Human inference: Strategies and shortcomings of social judgment_. Englewood Cliffs: Prentice-Hall, 1980.

Payne, J.W. Task complexity and contingent processing in decision making: An information search and protocol analysis. _Organizational Behavior and Human Performance_, 1976, _16_, 366-387.

Peterson, C. Recognition of noncontingency. _Journal of Personality and Social Psychology_, 1980, _38_, 727-734.

Schustack, M.W., & Sternberg, R.J. Evaluation of evidence in causal inference. _Journal of Experimental Psychology: General_, 1981, _110_, 101-120.

Sedlak, A.J., & Kurtz, S.T. A review of children's use of causal inference principles. _Child Development_, 1981, _52_, 759-784.

Seligman, M.E.P. On the generality of the laws of learning. _Psychological Review_, 1970, _77_, 406-418.

Seligman, M.E.P. _Helplessness_. San Francisco: Freeman, 1975.

Shapiro, A.K. A contribution to a history of the placebo effect. Behavioral Science, 1960, 5, 109-135.

Shultz, T.R., & Ravinsky, F.B. Similarity as a principle of causal inference. Child Development, 1977, 48, 1552-1558.

Shweder, R.A. Likeness and likelihood in everyday thought: Magical thinking in judgments about personality. Current Anthropology, 1977, 18, 637-658.

Siegler, R. S. The effects of simple necessity and sufficiency relationships on children's causal inferences. Child Development, 1976, 47, 1058-1063.

Siegler, R.S., & Liebert, R.M. Effects of contiguity, regularity, and age on children's causal inferences. Developmental Psychology, 1974, 10, 574-579.

Simon, H.A. Spurious correlation: A causal interpretation. Journal of the American Statistical Association, 1954, 49, 467-479.

Skinner, B.F. The phylogeny and ontogeny of behavior. Science, 1966, 153, 1205-1213.

Strawson, P.F. Introduction to logical theory. London: Methuen, 1952.

Toda, M. Causality, conditional probability and control. In A. Aykac & C. Brumat (Eds.), New developments in the application of Bayesian methods. Amsterdam: North Holland, 1977.

Tversky, A. Features of similarity. Psychologial Review, 1977, 84, 327-352.

Tversky, A., & Kahneman, D. Causal schemas in judgments under uncertainty. In M. Fishbein (Ed.), Progress in social psychology, Vol. 1. Hillsdale, NJ: Erlbaum, 1980.

Zajonc, R.B. Feeling and thinking: Preferences need no inferences. American Psychologist, 1980, 35 151-175.

TABLE 1

Effects of Shifts in the Causal Background

		Background (B)			
		Force of hammer	Defect in glass	Other explanations	
Background (C)	Force of hammer	23	0	0	23
	Defect in glass	32	2	2	36
	Other explanations	5	0	2	7
		60	2	4	66

Note: One subject responded to only one version of the stimulus and is therefore excluded from the analysis.

TABLE 2

Empirical Evidence Relating to Conflict between Contiguity and Perceived Similarity

		Contiguity: High	Contiguity: Low
ψ(Similarity)	High	Classical conditioning Michotte (1946)	Garcia *et al* (1968; 1972) Shultz & Ravinsky (1977)
	Low	Garcia *et al* (1968; 1972) Michotte (1946) Nisbett & Ross (1980) Shultz & Ravinsky (1977) Seligman (1970)	No relation

TABLE 3

Data Matrix for Hypothetical
Intercourse-Pregnancy Experiment

		Pregnancy: Yes	Pregnancy: No	
Intercourse	Yes	20	80	100
	No	5	95	100
		25	175	200

TABLE 4

Schematic Representation of the Theoretical Model

Level 4:	$S_n(Y,X_j \mid B) = F\left[s(Y',X_j \mid B) - \sum_{k=1}^{N} s(Y,X_k \mid B)\right]$	
Level 3:	$s(Y,X_j \mid B) = f(\text{Data, Content})$	
Level 2:	$\text{Data} = g_1\left[\Psi(r_{xy}), \text{robustness}\right]$	$\text{Content} = g_2\left[\text{temporal order, contiguity, } \Psi(\text{similarity})\right]$
Level 1:	$\Psi(r_{xy}) = h_1\begin{cases}(X\cap Y), (\bar{X}\cap\bar{Y}), (X\cap\bar{Y}), (\bar{X}\cap Y)\\ \text{rate, direction, level, noise}\end{cases}$	$\text{Contiguity} = h_2(\text{time, space})$ $\Psi(\text{Similarity}) = h_3(\text{common, distinctive elements})$

TABLE 5

Analysis of Variance for Experiment 1

Source of variation	df	MS	F	p <
Between subjects	31			
Groups	7	820.78	< 1	n.s.
Subjects within groups	24	1,687.50		
Within subjects	224			
Ψ(similarity) (A)	1	12,762.53	29.45	.01
Contiguity (B)	1	1,410.94	3.26	n.s.
Ψ(validity) (C)	1	41,692.53	96.21	.01
A x C	1	2,697.50	6.22	.05
B x C		1561.09	1.30	n.s.
Scenarios (D)	7	2,145.56	4.95	.05
D x A	7	1,397.28	3.22	.01
D x C	7	1,040.65	2.40	.05
D x A x C	7	620.10	1.43	n.s.
Error (within)	191	433.37		

TABLE 6

Mean Ratings of Causal Strength by Perceived Validity and Similarity: Experiment 1

		ψ(validity) High	ψ(validity) Low	
ψ(similarity)	High	58	26	42
	Low	37	19	27
		47	22	34

TABLE 7

Analysis of Variance for Experiment 2

Sources of variation		df	MS	F	p <
Between subjects		31			
Groups		7	1059.98	1.25	n.s
Subjects within groups		24	848.84		
Within subjects		96			
ψ(similarity)	(A)	1	5859.03	11.91	.01
Contiguity	(B)	1	2538.28	5.16	.03
ψ(validity)	(C)	1	9214.03	18.97	.01
A x C		1	39.06	< 1	n.s
B x C		1	342.25	< 1	n.s
Scenarios	(D)	3	2072.73	4.21	.01
Error (within)		88	492.15		

TABLE 8

Analysis of Variance for Experiment 3

Source of Variation	df	MS	F	p <
Between subjects	75			
Strength of alternative (A)	1	11,515.32	46.11	.001
Subjects with groups	74	249.72		
Within subjects	76			
Scenarios (B)	1	819.80	3.40	n.s
A x B	1	394.90	1.64	n.s.
B x subjects within groups	74	241.01		

TECHNICAL REPORTS DISTRIBUTION LIST

Engineering Psychology Programs
Code 442
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217 (2 cys)

Mathematics Research Programs
Code 411 MA
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

Statistics and Probability Program
Code 411 - S&P
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

Information Systems Program
Code 433
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

CDR K. Hull
Code 230
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

Physiology & Neuro Biology Programs
Code 441
Office of Naval Research
800 North Quincy Street
Arlington, VA 22217

Commanding Officer
ONR Eastern/Central Regional Office
ATTN: Dr. J. Lester
Building 114, Section D
666 Summer Street
Boston, MA 02210

Commanding Officer
ONR Western Regional Office
ATTN: Dr. E. Gloye
1030 East Green Street
Pasadena, CA 91106

Director
Naval Research Laboratory
Technical Information Division
Code 2627
Washington, DC 20375

Dr. Michael Melich
Communications Sciences Division
Code 7500
Naval Research Laboratory
Washington, DC 20375

Dr. Robert G. Smith
Office of the Chief of Naval
Operations, OP987H
Personnel Logistics Plans
Washington, DC 20350

Human Factors Department
Code N215
Naval Training Equipment Center
Orlando, FL 32813

Dr. Alfred F. Smode
Training Analysis and Evaluation Group
Naval Training Equipment Center
Code N-00T
Orlando, FL 32813

Dr. Jerry Lamb
Combat Control Systems
Naval Underwater Systems Center
Newport, RI 02840

Dr. Gary Poock
Operations Research Department
Naval Postgraduate School
Monterey, CA 93940

Mr. Warren Lewis
Human Engineering Branch
Code 8231
Naval Ocean Systems Center
San Diego, CA 92152

Dr. A. L. Slafkosky
Scientific Advisor
Commandant of the Marine Corps
Code RD-1
Washington, DC 20380

Naval Material Command
NAVMAT 0722 - Room 508
800 North Quincy Street
Arlington, VA 22217

Commander
Naval Air Systems Command
Human Factors Program
NAVAIR 340F
Washington, DC 20361

Mr. Fred M. Griffee
Command, Control & Communications
Code D 100A, Development Center
Marine Corps Development & Education Command
Quantico, VA 22134

CDR Robert Biersner
Naval Medical R & D Command
Code 44
Naval Medical Center
Bethesda, MD 20014

Dr. Arthur Bachrach
Behavioral Sciences Department
Naval Medical Research Institute
Bethesda, MD 20014

CDR Thomas Berghage
Naval Health Research Center
San Diego, CA 92152

Dr. George Moeller
Human Factors Engineering Branch
Submarine Medical Research Lab
Naval Submarine Base
Groton, CT 06340

Head
Aerospace Psychology Department
Code L5
Naval Aerospace Medical Research Lab
Pensacola, FL 32508

Dr. James McGrath
CINCLANT FLT HQS
Code 04E1
Norfolk, VA 23511

Dr. Robert Blanchard
Navy Personnel Research and
Development Center
Command and Support Systems
San Diego, CA 92152

CDR J. Funaro
Human Factors Engineering Division
Naval Air Development Center
Warminister, PA 18974

Dr. Julie Hopson
Human Factors Engineering Division
Naval Air Development
Warminster, PA 18974

Mr. Jeffrey Grossman
Human Factors Branch
Code 3152
Naval Weapons Center
China Lake, CA 93555

Human Factors Engineering Branch
Code 1226
Pacific Missile Test Center
Point Mugu, CA 93042

CDR C. Hutchins
Code 55MP
Naval Postgraduate School
Monterey, CA 93940

Technical Director
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

Director, Organizations and Systems
Research Laboratory
U.S. Army Research Institute
5001 Eisenhower Avenue
Alexandria, VA 22333

U.S. Air Force Office of Scientific Researcl
Life Sciences Directorate, NL
Bolling Air Force Base
Washington, DC 20332

Chief, Systems Engineering Branch
Human Engineering Division
USAF AMRL/HES
Wright-Patterson AFB, OH 45433

Dr. Earl Alluisi
Chief Scientist
AFHRL/CCN
Brooks AFB, TX 78235

Dr. Kenneth Gardner
Applied Psychology Unit
Admiralty Marine Technology Establishment
Teddington, Middlesex, ENGLAND TW11 OLN

Director, Human Factors Wing
Defence & Civil Institute of
Environmental Medicine
Post Office Box 2000
Downsview, Ontario M3M 3B9, CANADA

Dr. A. D. Baddeley, Director
Applied Research Council
15 Chaucer Road
Cambridge, ENGLAND CB2 2EF

Defense Technical Information Center
Cameron Station, Building 5
Alexandria, VA 22314 (12 cys)

Dr. Judith Daly
Systems Sciences Office
Defense Advanced Research Projects Agency
1400 Wilson Boulevard
Arlington, VA 22209

Dr. Robert R. Mackie
Human Factors Research, Inc.
5775 Dawson Avenue
Goleta, CA 93017

Dr. Jesse Orlansky
Institute for Defense Analyses
400 Army-Navy Drive
Arlington, VA 22202

Dr. T. B. Sheridan
Department of Mechanical Engineering
Massachusetts Institute of Technology
Cambridge, MA 02139

Dr. Paul Slovic
Decision Research
1201 Oak Street
Eugene, OR 97401

Dr. Harry Snyder
Department of Industrial Engineering
Virginia Polytechnic Institute
State University
Blacksburg, VA 24061

Dr. Amos Tversky
Department of Psychology
Stanford University
Stanford, CA 94305

Dr. Robert T. Hennessy
NAS - National Research Council
2101 Constitution Avenue, N.W.
Washington, DC 20418

Dr. M. G. Samet
Perceptronics, Inc.
6271 Variel Avenue
Woodland Hills, CA 91364

Dr. Robert Williges
Human Factors Laboratory
Virginia Polytechnic Institute
and State University
130 Whittemore Hall
Blacksburg, VA 24061

Dr. Ward Edwards, Director
Social Science Research Institute
University of Southern California
Los Angeles, CA 90007

Dr. Kenneth Hammond
Institute of Behavioral Science
University of Colorado, Room 201
Boulder, CO 80309

Dr. James H. Howard, Jr.
Department of Psychology
Catholic University
Washington, DC 20064

Dr. William Howell
Department of Psychology
Rice University
Houston, TX 77001

Dr. Christopher Wickens
University of Illinois
Department of Psychology
Urbana, IL 61801

Dr. Richard W. Pew
Information Sciences Division
Bolt Beranek & Newman, Inc.
50 Moulton Street
Cambridge, MA 02238

Dr. John Payne
Duke University
Graduate School of Business Administration
Durham, NC 27706

Dr. Baruch Fischhoff
Decision Research
1201 Oak Street
Eugene, OR 97401

Dr. Andrew P. Sage
University of Virginia
School of Engineering and Applied Science
Charlottesville, VA 22901

Dr. Lola Lopes
Department of Psychology
University of Wisconsin
Madison, WI 53706

Mr. Joseph G. Wohl
Alphatech, Inc.
3 New England Industrial Park
Burlington, MA 01803

Dr. Rex Brown
Decision Science Consortium
7700 Leesburg Pike, Suite 721
Falls Church, VA 22045

Dr. Wayne Zachary
Analytics, Inc.
2500 Maryland Road
Willow Grove, PA 19090

Dr. Marvin Cohen
Decision Science Consortium
7700 Leesburg Pike, Suite 721
Falls Church, VA 22043

LMED

8